

ExFuse: Enhancing Feature Fusion for Semantic Segmentation

Zhenli Zhang¹, Xiangyu Zhang², Chao Peng², Dazhi Cheng³, Jian Sun²

¹Fudan University, ²Megvii Inc. (Face++), ³Beijing Institute of Technology
zhenlizhang14@fudan.edu.cn, {zhangxiangyu, pengchao}@megvii.com,
1120152020@bit.edu.com, sunjian@megvii.com

ECCV 2018

目的:

本文主要研究了语义分割网络中的高级和低级特征的融合效率，像 Unet，当特征层级或空间分辨率存在较大差距时，特征的融合效率较低。文章提出了几种特征融合方式，为了提高低级特征的语义信息，在高级特征中嵌入更多空间信息。

方法:

低级特征中引入更多语义信息:

layer rearrangement :

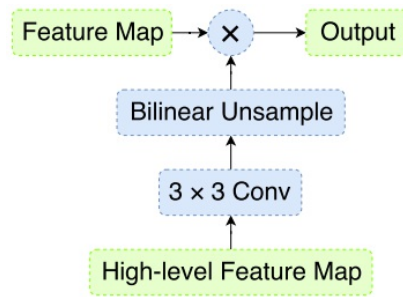
ResNeXt 网络结构中，各级的网络包含的残差单元个数为{3,4,23,3}。为了提高底层特征的语义性，让低层的两级网络拥有的层数更多。因此将残差单元个数重排为{8,8,9,8}。mIoU 提升 0.8%。

semantic supervision :

多层语义监督是指把辅助监督直接指派到编码器的早期阶段，它主要专注于完善低级特征的质量，而不是提升模型本身。mIoU 提升 1.1%。

semantic embedding branch :

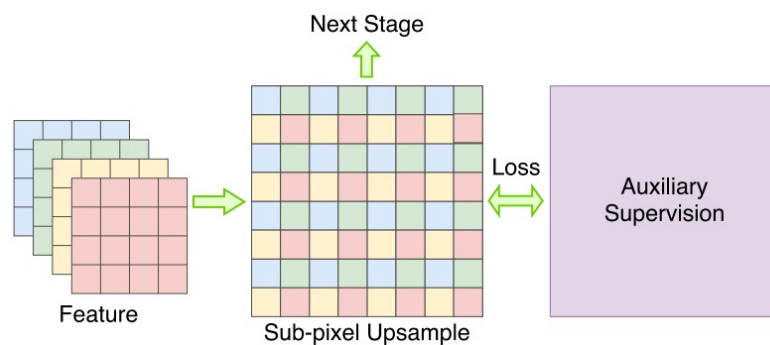
从高级特征中包含更多的语义信息来指导分辨率融合。mIoU 提升 0.7%。



在高级特征中嵌入更多空间信息：

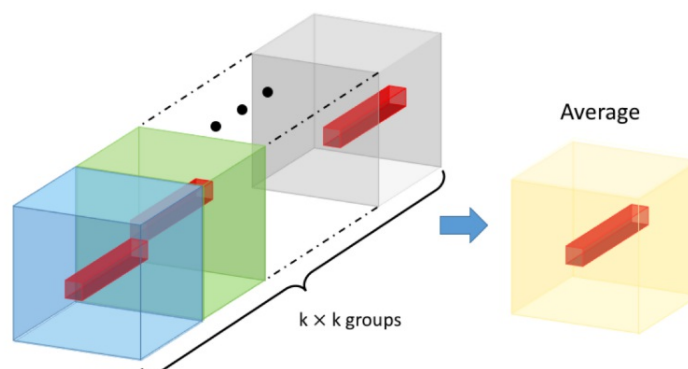
Explicit Channel Resolution Embedding :

文中分析了转置卷积加辅助损失性能没有提升的原因：辅助损失是将高分辨率信息显式地嵌入到特征图通道中，然而转置卷积包含权值让嵌入变得隐式。使用 Sub-pixel Upsample 与转置卷积，加上辅助损失可以解决这个问题。mIoU 提升 0.5%。



Densely Adjacent Prediction :

将空间信息编码到通道中，用在最后预测阶段，让每个位置的预测取决于周围邻域。作用域看 $k \times k$, 先扩展通道到 $K \times K$ 倍, 然后将每一组中的邻域特征平均。



网络结构：

